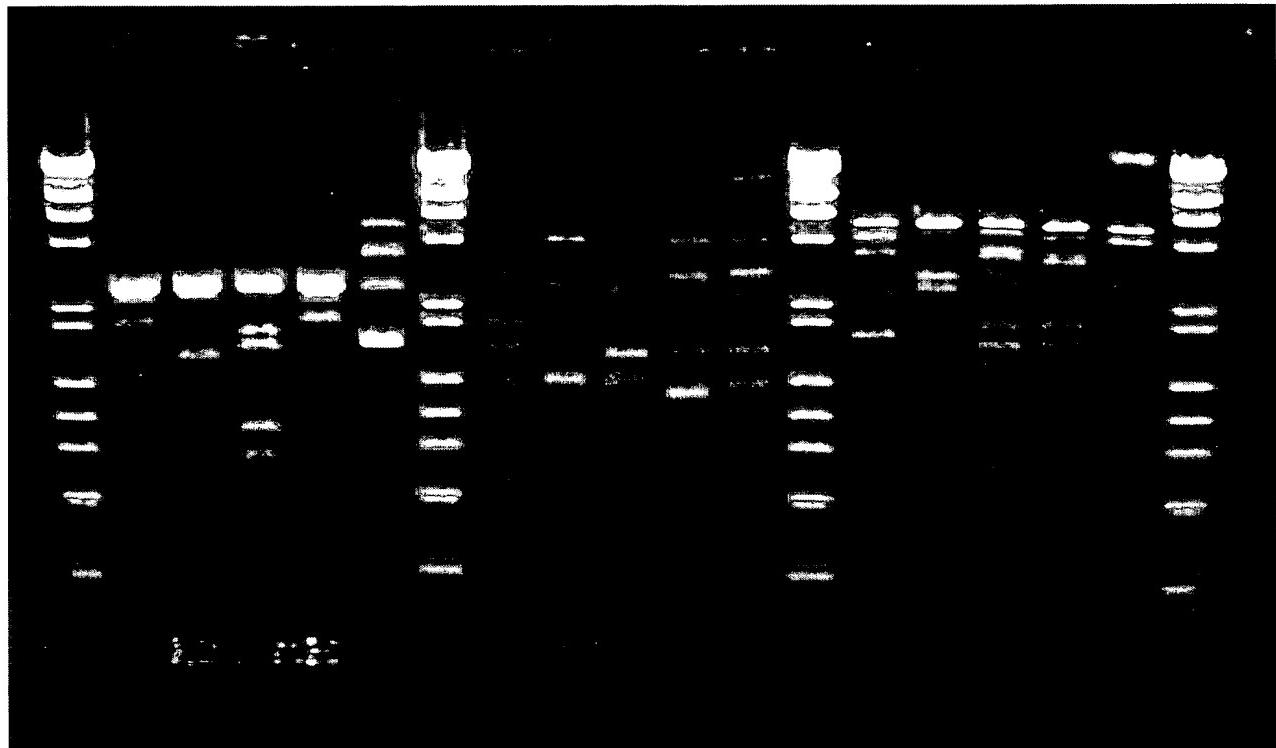


CstMI Figure 1 - Agarose gel showing CstMI cleavage of  
lambda, T7, phiX174, pBR322 and pUC19 DNAs.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19  
20



CstMI Figure 2 – DNA sequence of the CstMI gene locus (SEQ ID NO:).

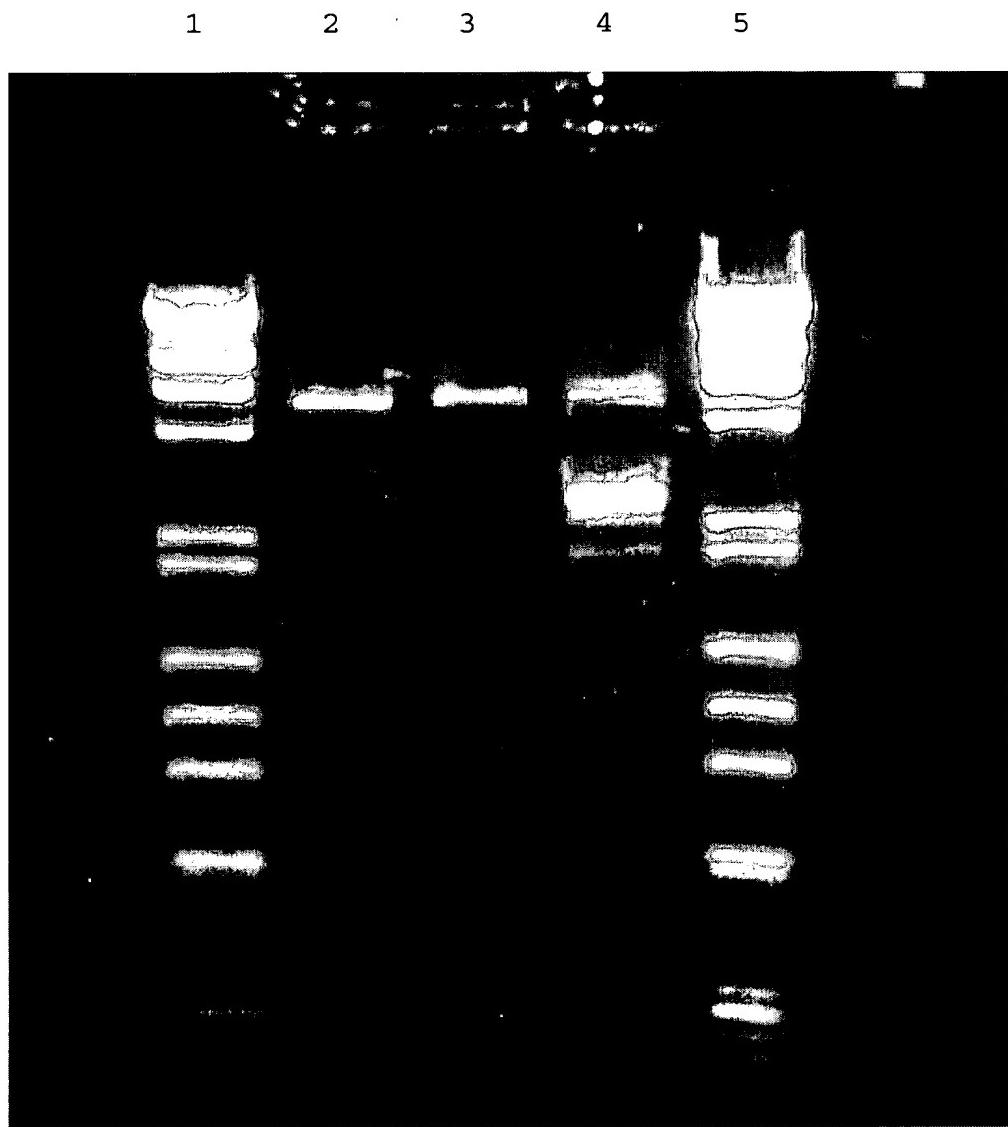
1 ATGGTTATGG CCCCTACGAC TGTTTTGAC CGCGCTACCA TTGCCACAA  
51 TCTCACCGAA TTCAAACCTCC GGTGGCTTGA CCGCATTAAAG CAATGGGAGG  
101 CGGAAAACCG ACCCGCAACC GAGTCGAGTC ACGACCAACA GTTCTGGGT  
151 GACCTGCTCG ACTGCTTCGG TGTCAACGCC CGCGACCTGT ACTTGTACCA  
201 ACGCAGCGCT AAACGCGCTT CGACGGGGCG CACCGGCAAG ATCGACATGT  
251 TTATGCCGGG CAAAGTCATA GGCGAGGCTA AGTCCCTCGG CGTCCCCTGC  
301 GATGATGCTT ATGCCCAAGC TTTGGATTAT TTGCTGGCG GTACTATCGC  
351 GAACTCGCAC ATGCCGGCCT ATGTTGTCTG CTCCAACCTTC GAGACCCCTGC  
401 GGGTTACCCG TCTTAACCGC ACCTATGTCTG GCGATAGCGC CGACTGGGAC  
451 ATTACATTCC CTTTAGCTGA GATTGACGAG CACATCGAAC AACTCGCTT  
501 TCTCGCCGAC TATGAAACCT CCGCCTACCG GGAGGAAGAA AAGGCTTCCC  
551 TGGAAAGCCTC TCGGTTAATG GTGGAGCTCT TCCCGGCCAT GAACGGCGAC  
601 GACGTGGACG AGGCAGTAGG CGATGACGCT CCCACCACGC CGGAGGAAGA  
651 AGACGAGCGC GTCATGCGCA CCTCTATCTA CCTCACCCGA ATCCTCTTCC  
701 TTCTCTTCGG CGACGACGCA GGACTCTGGG ATACCCCGCA TTTGTTTGC  
751 GACTTTGTGC GCAATGAAAC CACCCCCAGAA TCGCTGGCC CGCAGCTCAA  
801 TGAGCTATTT AGCGTGCTTA ATACCGCCCC GGAAAAGCGG CCTAAGCGTT  
851 TGCCATCAAC GTTGGCGAAG TTTCCTTATG TCAATGGTGC CCTATTGCT  
901 GAACCGTTGG CCTCGGAGTA CTTCGACTAC CAGATGCGCG AAGCATTGCT  
951 TGCTGCCTGC GACTTCGACT GGTGACCAT TGACGTCTCC GTCTTGGTT  
1001 CGTTGTTCCA ATTGGTAAA TCGAAGGAAG CGCGCCGCAG CGACGGCGAA  
1051 CACTACACGT CTAAGGCCAA CATCATGAAG ACCATCGGCC CGCTGTTTT  
1101 GGACGAGCTG AGGGCTGAGG CCGATAAGTT GGTGTCTTCT CCGTCGACGT  
1151 CGGTGGCCGC ATTAGAGCGC TTCCGCGACT CCCTGTCTGA GCTGGTATTC  
1201 GCTGATATGG CTTGTGGTTC TGAAAATTC CTGCTTCTGG CGTATCGGG  
1251 GTTGCGCCGG ATTGAAACCG ACATCATTGT CGCTATACGC CAGCGCCGCG  
1301 GTGAAACGGG CATGTCGTTG AATATTGAGT GGGAGCAGAA ACTGTCCATT  
1351 GGGCAGTTCT ACGGCATTGA GCTGAATTGG TGGCCTGCCA AGATTGCTGA  
1401 GACTGCCATG TTCCTAGTTG ACCATCAGGC CAACAAGGAG CTTGCCAACG  
1451 CTGTGGGTAG GCCTCCGGAG CGGTTGCCGA TTAAGATTAC CGCGCACATT  
1501 GTGCACGGCA ATGCCCTGCA GCTTGATTGG GCAGACATAC TCTCGGCTTC

1551 TGCCGCCAAG ACGTATATCT TCGGTAACCC GCCGTTTTG GGGCATGC  
1601 CGAGAACTGC TGAACAAGCT CAAGAACTCC GAGACTTGTG GGGCACTAAG  
1651 GACATTTCAC GCTTGGACTA CGTCACCGGC TGGCATGCAA AGTGCTTGG  
1701 TTTCTTTAAG TCCCAGAGGG GTGCGTTTGC GTTTGTCACC ACCAATTCAA  
1751 TTACTCAAGG TGATCAAGTT CCACGGCTAT TTGGGCCTAT CTTCAAAGCA  
1801 GGGTGGCGTA TTCGTTTCGC TCACCGCACG TTTGCGTGGG ACTCTGAAGC  
1851 ACCCGGTAAA GCTGCTGTT ACTGCGTCAT TGTTGGCTTC GATAAGGAGA  
1901 GTCAACCACG TCCACGTCTG TGGGATTATC CCGATGTAAA GGGCGAGCCA  
1951 GTCTCAGTGG AAGTAGGCCA GTCCATTAAT GCCTATTTAG TAGACGGCCC  
2001 TAATGTTCTT GTCGATAAAAT CCCGGCATCC TATTCGTCG GAAATATCGC  
2051 CCGCAACTTT TGGAAATATG GCGCGAGATG GCGGCAACCT TCTAGTTGAG  
2101 GTCGACGAAT ACGACGAGGT TATGAGTGAC CCCGTAGCGG CAAAGTATGT  
2151 TCGCCCTTTC CGGGTAGTC GAGAGCTAAT GAACGGCTTA GATCGGTGGT  
2201 GTCTATGGCT TGTAGATGTA GCACCGTCAG ACATTGCCA GAGTCCGGTT  
2251 CTGAAAAAGC GTCTAGAACG GTTAAAGTCT TTTCGAGCCG ACAGTAAAGC  
2301 GGCAAGTACA CGGAAAATGG CTGAAACTCC GCACTTATTG GGCCAGCGG  
2351 CGCAACCGGA TACTGATTAC CTTTGCCTGC CGAAGGTAGT AAGCGAACGC  
2401 CGCTCGTATT TCACCGTACA AAGGTATCCA TCAAACGTAA TCGCTTCTGA  
2451 CCTAGTATTC CATGCTCAAG ATCCAGACGG CCTGATGTTT GCGCTAGCGT  
2501 CGTCGTCGAT GTTCATTACG TGGCAGAAAA GCATCGGAGG ACGACTCAAG  
2551 TCTGATCTCC GTTTGCTAA CACTTGACG TGGAATACTT TCCCAGTGCC  
2601 AGAACTCGAC GAGAAGACGC GGCAGCGAAT TATTAAAGCG GGCAAGAAGG  
2651 TGCTCGACGC CCGCGCGCTG CACCCAGAAC GCTCGCTGGC CGAGCACTAC  
2701 AACCCACTCG CGATGGCACC GGAACTCATC AAAGCGCATG ATGCGCTCGA  
2751 CCGCGAGGTG GATAAAGCGT TTGGCGCGCC ACGAAAGCTG ACAACTGTT  
2801 GGCAGCGCCA GGAGCTATTG TTTGCCAATT ACGAAAAACT CATCTCACAC  
2851 CAGCCCTAG

CstMI Figure 3: Amino acid sequence of the CstMI gene locus (SEQ ID NO:).

1 MVMAPTTVFD RATIRHNLTE FKLRWLDRIK QWEAENRPAT ESSHDDQQFWG  
51 DLLDCFGVNA RDLYLYQRSA KRASTGRTGK IDMFMMPGKVI GEAKGSLGVPL  
101 DDAYAQALDY LLGGTIANSH MPAYVVCSNF ETLRVTRLNR TYVGDSADWD  
151 ITFPLAEIDE HIEQLAFLAD YETSAYREEE KASLEASRLM VELFRAMNGD  
201 DVDEAVGDDA PTTPEEEDER VMRTSILTR ILFLLFGDDA GLWDTPHLFA  
251 DFVRNETTPE SLGPQLNELF SVLNNTAPEKR PKRLPSTLAK FPYVNGALFA  
301 EPLASEYFDY QMREALLAAC DFDWSTIDVS VFGSLFQLVK SKEARRSDGE  
351 HYTSKANIMK TIGPLFLDEL RAEADKLVSS PSTSVAALER FRDSLSELVF  
401 ADMACGSGNF LLLAYRELRR IETDIIVAIR QRRGETGMSL NIEWEQKLSI  
451 GQFYGIELNW WPAKIAETAM FLVDHQANKE LANAVGRPPE RLPIKITAH  
501 VHGNALQLDW ADILSASAATK TYIFGNPPFL GHATRTAEQA QELRDLWGTK  
551 DISRLDYVTG WHAKCLDFFK SREGRFAFVT TNSITQGDQV PRLFGPIFK  
601 GWRIRFAHRT FAWDSEAPGK AAVHCIVVGF DKESQPRPRL WDYPDVKGEP  
651 VSVEVGQSIN AYLVDGPNVL VDKSRHPISS EISPATFGNM ARDGGNLLVE  
701 VDEYDEVMSD PVAAKYVRPF RGSRELMNGL DRWCLWLVDV APSDIAQSPV  
751 LKKRLEAVKS FRADSKAAST RKMAETPHLF GQRSPDQTDY LCLPKVVSER  
801 RSYFTVQRYP SNVIASDLVF HAQDPDGLMF ALASSSMFIT WQKSIGGRLK  
851 SDLRFANTLT WNTFPVPELD EKTRQRIIKA GKKVLDARAL HPERSLAEHY  
901 NPLAMAPELI KAHDALDREV DKAFGAPRKL TTVRQRQELL FANYEKLISH  
951 QP

Figure 4 - Agarose gel showing CstMI protection of pTBCstMI.3 DNA and cleavage of unmodified DNA substrate.



CstMI Figure 5 : Determination of the CstMI cleavage site.

Figure 5A: Location of cleavage on 5'-AAGGAG-3' strand.

pUC19-Adeno2BC4 DNA was cut with CstMI producing ends as indicated by the arrows:

5'-. . . CGAACCCAGGTGTGCGACG↓TCAGACAACGGGGAGCGCTCCTTTTG..-  
3'  
(SEQ ID NO:3)  
3'-. . . GCTTGGGTCCACACGCT↑GCAGTCTGTTGCCCTCGCGAGGAAAAC..-  
5'

The resulting cleaved DNA:

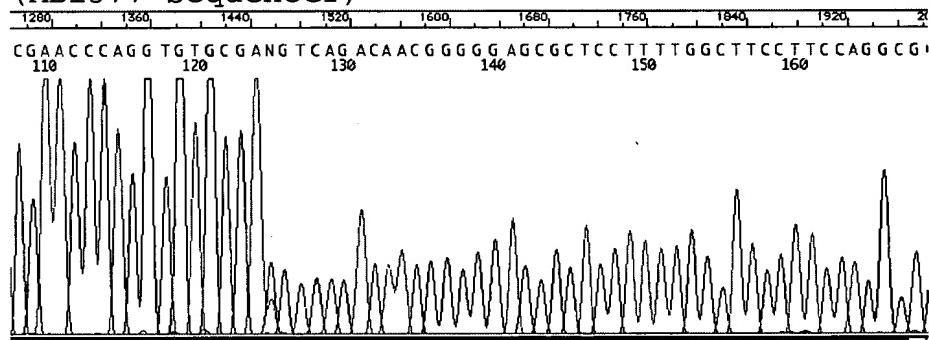
5'-. . . CGAACCCAGGTGTGCGACG-3' (SEQ ID NO:4)  
3'-. . . GCTTGGGTCCACACGCT-5'

The template strand for dideoxy DNA sequencing extension:  
3'-. . . GGGTCCACACGCT-5'

The primer (NEB1224) is annealed and extended through the CstMI site. When the reaction reaches the end of the molecule the Taq polymerase adds an extra A base.

5'-PRIMER->. . . CGAACCCAGGTGTGCGA(A)-3' (SEQ ID NO:5)  
3'-. . . . . . . GCTTGGGTCCACACGCT-(N20-GAGGAA)-5'

Sequencing Profile of CstMI cut pUC19-Adeno2BC4 DNA (ABI377 Sequencer)



CstMI Figure 5: Determination of the CstMI cleavage site.

Figure 5B: Location of cleavage on 5'-CTCCTT-3' strand.

pBR322 DNA was cut with CstMI, yielding ends indicated by the arrows:

5' -.. TGCATGC **AAGGAG** ATGGCGCCCAACAGTCCCCC ↓ GGCCACGGGGCC .. -  
3'  
(SEQ ID NO:6)

3' -.. ACGTACG **TTCCTC** TACCGCGGGTTGTCAGGG ↑ GGCCGGTGCCCCGG .. -  
5'

The resulting cleaved DNA:

5' -.. TGCATGC **AAGGAG** ATGGCGCCCAACAGTCCCC -3'  
(SEQ ID NO:7)

3' -.. ACGTACG **TTCCTC** TACCGCGGGTTGTCAGGG -5'

The template strand for dideoxy DNA sequencing extension:

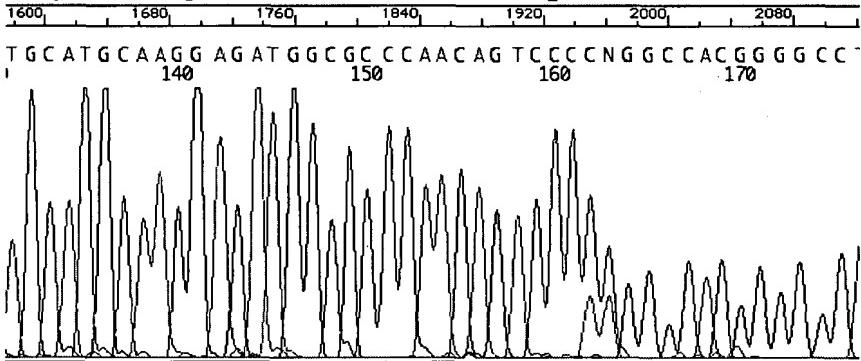
3' -.. ACGTACG **TTCCTC** TACCGCGGGTTGTCAGGG -5'

The primer (NEB1242) is annealed and extended through the CstMI site. When the reaction reaches the end of the molecule the Taq polymerase adds an extra A base.

5' -PRIMER->. TGCATGC **AAGGAG** ATGGCGCCCAACAGTCCC (A) -3'  
(SEQ ID NO:8)

3' -.. . . . ACGTACG **TTCCTC** TACCGCGGGTTGTCAGGG -5'

Sequencing Profile of CstMI pBR322 DNA (ABI377 Sequencer)



CstMI Figure 6: Sequence alignment of CstMI and MmeI amino acid sequences.

Symbol comparison table: /gcb/bin/gcgcore/data/rundata/blosum62.cmp  
CompCheck: 1102

Gap Weight: 8 Average Match: 2.778  
Length Weight: 2 Average Mismatch: -2.248

Quality: 1548 Length: 942  
Ratio: 1.718 Gaps: 19

Percent Similarity: 51.009 Percent Identity: 39.574

Match display thresholds for the alignment(s):

| = IDENTITY  
: = 2  
. = 1

CstMI.pep x MmeI.pep June 20, 2003 11:45 ..

20 EFKLRWLDRIKQWEAENRPATESSHQQFWGDLDCFGVNARDLYLYQRS 69  
| : : :: | . || . . | . | | : | : . . : . : .  
7 EIRRKAIEFSKRWE...DASDENSAQAKPFLIDFFEVFGITNKRVATFEHA 53

70 AKRASTGR...TGKIDMFMPGKVIGEAKSLGVPLDDAYAQALDYLLGGT 115  
| : . | : | : | . : | || | || | || | || | || | |  
54 VKKFAKAHKEQSRRGFVDLFWPGILLIEMKSRGKDLDKAYDQALDYFSG.. 101

116 IANSHMPAYVVCSNFETLRVTRLNRRTYVGDSADWDITFPPLAEIDEHIEQL 165  
|| : | | . . : | . | | : | : | | : | : . : .  
102 IAERDLPRYVLVCDFQRFRLTDL...ITKES....VEFLLKDLQNVRSF 144

166 AFLADYETSAYREEEKASLEASRLMVELFRAMNGDDVDEAVGDDAPTTPE 215  
| : | : | : : . . : | . | | : | . | : | :  
145 GFIAGYQTQVIKPQDPINIKAAERMGKL.....HDTLKLVGYEGHA... 185

216 EEDERVMRITSIYLTRILFLFGDDAGLWDTPHLFADFVRNETTPE..SLG 263  
| : | : | : | : | : . : | : | : | : | : | : |  
186 .....LELYLVRLLFCLFAEDTTIFE.KSLFQEYIETKTLEDGSGLA 226

264 PQLNELFSVLNTAPEKRPKRLPSTLAKFPYVNGALFAEPLASEYFDYQMR 313  
| : | | | | : | : | | | | : | | | | | | | | |  
227 HHINTLFYVLNTPEQKRLKNLDEHLAAFPYINGKLFEPLPPAQFDKAMR 276

314 EALLAACDFDWSTIDVSVFGSLFQLVKSKEARRSDGEHYTSKANIMKTIG 363  
| | | | | | | : | : | | | | : | | | | | | | |  
277 EALLDLCSLDWSRISPAIFGSLFQSIMDAKKRRNLGAHYTSEANILKLIK 326

364 PLFLDELRAEADKLVSSPSTSVAALERFRDSLSELVFADMACGSGNLLL 413  
| | | | | | : | . . . . | | | | | | | | | | | |  
327 PLFLDELWVEFEKVKNNKNKLLA...FHKKLRLGLTFDPACGCGNFLVI 372

414 AYRELRIETDIIVAIRQRRGETGMSLNIEWEQKLSIGQFYGIELNWPA 463  
| | | | : | : : : | | : | . | | . : | : | : | : |  
373 TYRELRLLEIEVLRGL.HRGQQ..QVL DIEHLIQINVDQFFGIEIEEFPA 419

464 KIAETAMFLVDHQANKELANAVGRPPERLPIKITAHIVHGNALQDWADI 513  
| . | : | : | | | | . . . | | : | : | | | | | | | |  
420 QIAQVALWTDHQMNMKISDEFGNYFARIPLKSTPHILNANALQIDWNDV 469

514 LSASAALKTYIFGNPPFLGHATRTAEQAQELRDLWGT.KDISRLDYVTGWH 562  
| | | : | | | | . | . . | | : | . | | | | | | | | : |

470 LEAKKC.CFILGNPPFVGKSKQTPGQKADLLSVFGNLKSASDLDLVAAWY 518  
563 AKCLDFFKSREG.RFAFVTTNSITQGDQVPRLFPGPIFKAGWRIRFAHRTF 611  
| : .. | |||.|||||||:|| . : | :| |||||  
519 PKAAHYIQTNANIRCAFVSTNSITQGEQVSLLWPLLLSLGIKINFAHRTF 568  
612 AWDSEAPGKAAVHCIVGFDKESQPRPRLWDYPDVKGEPVSVEVGQSINA 661  
.|| .|| | |||||:|| . : ::| : |||..: ..||  
569 SWTNEASGVAAVHCVIIGFGLKDSDEKIIYEYESINGEPLAIK.AKNINP 617  
662 YLVDGPNVLVDKSRHPISSEISPATFGNMARDGGNLLVEVDEYDE.VMSD 710  
|| | .|: | . |.: | :| | ||| | :| .. : ..  
618 YLRDGVDVIACKRQQPI.SKLPSPMRYGNKPTDDGNFLFTDEEKNQFITNE 666  
711 PVAAKYVRPFRGSRELMNGLDRWCLWLVDVAPS DIAQSPV LKKRLEAVKS 760  
| . | | | | . | | ||||| |:| | .. |: . |.  
667 PSSEKYFRRFVGGDEFINNTSRWCLWLDGADISEIRAMPLVLARIKKVQE 716  
761 FRADSKA ASTRKMAETPHLFGQRSQPDTDYLC LPKVVSERRSYFTVQRYP 810  
|| | | | . | | | | | ||||| | :| . | | | : :  
717 FRLKSSAKPTRQSASTPMKFFYISQPD TDYLLIPETSSEN RQFIPIGFVD 766  
811 SNVIASDLVFHAQDPDGLMFALASSSMFITWQKSIGGRLKSDLRFANTLT 860  
| |||. | : | . | | | | | : :| | | | | : .. |.  
767 RNVIISSNATYHIPS A EPLIFGLLSSTMHNCWMRNV GGRLESRYRYSASLV 816  
861 WNTFPVPELDEKTRQRIIKAGKKVLDARALH PERSLAEHYNPLAMAPELI 910  
:| | | | : . | | . | : | | . | : | | | | | | | | | :  
817 YNTFPWIQPNEKQSKAIEEAFAILKARSNYPNESLAGLYDPKTMPSELL 866  
911 KAHDALDREV DKAFGAPRKLT TVRQRQELLFANYEKLISHQP 952  
| | | | : | | : | | : | | | | | | | | | | | | | |  
867 KAHQKLDKAVDSVYGF KGPNTEI.ARIAFLFETYQKMTSLLP 907

FIGURE 7

